Document Overview

This document details the Phase 5 design of the Medi-Cal Management Information System and Decision Support System (MIS/DSS). The MIS/DSS transforms raw provider, financial, eligibility, prescription drug, managed care encounter and fee-for-service claims data into an integrated, knowledge-based information system to support key management and policy decisions of the Medi-Cal program.

This document contains the following sections:

- Introduction — This section includes the Document Overview, a Summary of Design Changes, and the System Design Walkthrough Action Items.
- Overviews — This section includes high-level overviews of the MIS/DSS System and the primary databases, as background for the detailed design specifications that follow. Specifically, this section includes the following overviews:
  - System
  - DataScan Database
  - Panorama View Database
  - Performance Measurement Workstation Database
- Splitter — This section contains the logic for the pre-convert of the RF-035 File.
- GeoCoding — This section contains the logic for applying latitude and longitude values to the Provider, Medical Claims, and Drug tables.

- **Data Enhancement Functional Specifications by Table**

  A separate tab and set of documents exists for each DB2 table. Beginning in Phase 3, MEDSTAT initiated a reformatting of all design specifications into functional specifications. The intention of this effort is to make the detailed design of the MIS/DSS components easier to understand for the majority of readers. The documents have been split into two pieces: a) a background document describing the key processes; and b) a field level specification (if applicable) denoting the characteristics and logic employed for a given field.

  The following DB2 tables are included:
  - Eligibility
  - Provider
  - Medical Claims (Inpatient and Outpatient Service Tables)
  - Drug
  - Case
  - Episodes
  - Capitation
  - Managed Care Financial

- **Product Customizations**

  This section outlines the customized design applied to the Panorama View, Panorama Briefing Book, Performance Measurement Workstation, and MyEureka! products.

- **Reference Maps**

  This section includes all of the reference maps used by the DB2 table conversion programs.

In the discussion of software products throughout this document please note the following: DataScan, Panorama View, Panorama Briefing Book, and Performance Measurement Workbook are registered trademarks of The MEDSTAT Group. MapInfo, SAS, MyEureka! are registered trademarks of MapInfo Corporation, the SAS Institute, and Information Advantage, respectively.

Medi-Cal Management Information
System and Decision Support System (MIS/DSS)

System and Database Overviews

Phase 5

March 23, 2000

Table of Contents

1. System Overview

1.1.1 Background

The California Department of Health Services (DHS) administers and oversees the implementation of the Medi-Cal healthcare program for Medicaid beneficiaries. Within the overall program there are a number of types of coverage including:

- Office visits for illness
- Outpatient hospitalization for accidental emergency and minor surgery
- Inpatient hospitalization
- Long term care
- Laboratory Services
- Prescription drugs
- Wellcare visits for babies and pregnant mothers (CHDP)
- Immunizations
- Mental health services (Short Doyle)

The recent trend toward local administration of State and Federal entitlement programs, along with the development of managed healthcare, has led to several new initiatives for the administration of the Medi-Cal program including:

- The State implemented Geographic Managed Care Initiative (GMC)
- County Organized HealthCare Systems (COHS)
- Two Plan Healthcare Models

As a result of the diverse coverage and methods of administration for the Medi-Cal program, it has become more difficult to access the data necessary to monitor the efficiency and effectiveness of the overall program. In an effort to resolve this concern, DHS has initiated the development of a Health Data Warehouse that serves to “centralize” data from all of the various programs regardless of the administrator. This data warehouse and associated applications in the MEDSTAT-supplied suite constitute the Medi-Cal Management Information System/Decision Support System (MIS/DSS).

The Health Data Warehouse not only centralizes data but also serves to make diverse data much more useful. This additional value is a result of:

- Enhancing the data using standard methodologies
- Ensuring the quality of the data
- Improving the quality of the data

1.1.2 Goals Of Data Enhancement

The main goal of data enhancement, of course, is to produce databases, which support the needs of the State. More specifically, data enhancement should do the following:

- Standardize disparate data types in such a way as to make all the data comparable at a high level.
    - *Example*: Service codes used for billing in California include ICD-9, CPT, HCPCS, local codes only valid in CA, and LA Waiver codes. The standardization process seeks to identify all common codes by their respective meaning and apply the same national CPT/HCPCS codes to them in the Performance Management Workstation (PMW) database, because the HEDIS reports created in PMW use criteria recognized only by these codes.
- At the same time, maintain meanings familiar to users of limited types of data, such as dental, medical FFS, mental health, etc.
    - *Example*: MEDSTAT categorizes services into the Federal Category of Service values used for HCFA reporting. This categorization is based on the criteria the DHS Medical Care Statistics Section currently uses in their data feed to HCFA.
- Support application-specific processing, such as DataScan® Case and Episode builds, so that the applications present meaningful and productive reporting and analyses.

  Examples:
    - The DataScan® Claims core fields MEMBERNO, PLACE and SVCTYP are determined largely by how values in these fields will affect Case build.
    - The Panorama View® dimensions drive the logic for their source fields in the DataScan® Claims, Drug and Eligibility tables.
- Provide for cross-product comparability, so that, for example, Panorama View reports can be supported by detailed analysis in the DataScan® database.
    - *Example*: Total net payments on the DataScan® Decision Support System (DSS) match Total Expenditures reported by the Panorama View dimensions.
- Support comparability to external benchmarks, including Standard State reports and MEDSTAT norms, for both commercial and Medicaid clients.

- *Example*: the logic for the DataScan® Claims and Drug custom field SVCCAT is driven by the desire to tie certain Panorama View® reports to 2082 reports.

1.1.3 The MIS/DSS Databases And Applications

The MIS/DSS consists of three databases: DataScan®, Panorama View®, and Performance Measurement Workstation® (PMW). Data for Panorama View and PMW is extracted from the DataScan® database. Each of these databases serve different purposes and support different applications. The DataScan® database contains all of the detailed information about the claims, eligibles, providers, and managed care plans. The Panorama View® database is a multi-dimensional summarization of information contained in DataScan® (e.g., claims aggregated by categories such as aid code and category of service). It summarizes all the records in the DataScan® database, but only for selected fields (called dimensions). The PMW database contains detail records extracted from the DataScan® database, but only those records that are relevant to HEDIS measures of healthcare quality.

There are seven applications in the MEDSTAT suite (see Figure 1). The MEDSTAT applications are intended to allow data to be accessed either in aggregate (Panorama View), in a report format (PMW and Briefing Book), or queried at a detailed level (DataScan® and MyEureka!). SAS allows more sophisticated power users the ability to utilize multivariate statistics to analyze the DataScan® data. The MapInfo product provides the capacity to perform geographic analysis of eligibles and providers. The diversity of these products allows the user to "mix or match" views and to analyze the data in any number of ways depending upon their needs, their depth of technical skill, and knowledge of the underlying data.

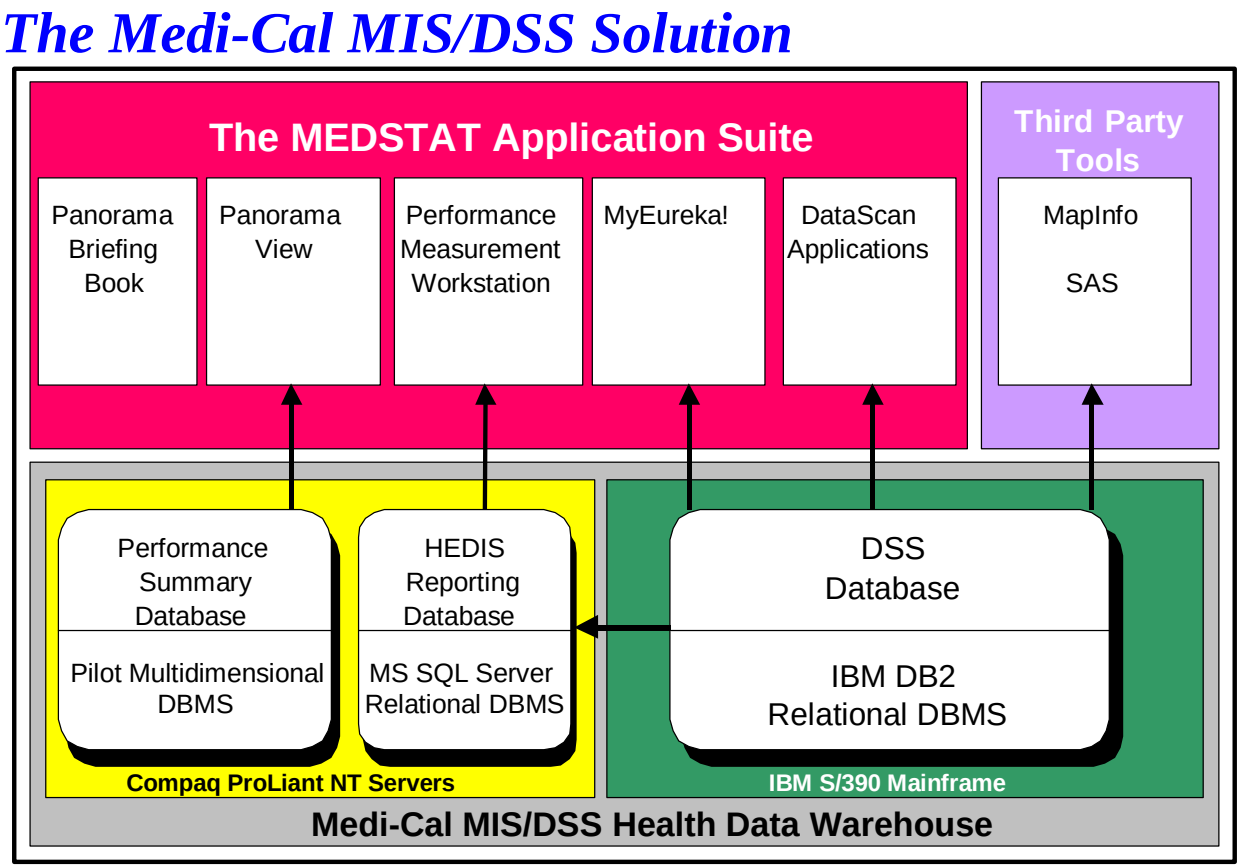

Figure 1. The MIS/DSS Application Suite

1.1.4 Database Construction Processes

The MIS/DSS databases are constructed via a conversion process that begins with the development of the DataScan® database. Programs are executed to perform data conversion (either via known patterns, tagging from other sources, or mappings from lookup tables). The conversion programs cover five primary areas:

- Medical and Prescription Drug claims
- Providers
- Eligibility
- Capitation Payments
- Managed care plan financial reports

Once the conversions occur, extract files are created. For Panorama View®, extracts of claims, providers, case days and capitation data are created and passed to the Panorama View build process. For PMW, a specific set of HEDIS-related claim and eligibility records are extracted and sent to the PMW build process. The extract process ensures that all products are working from the same baseline set of Medi-Cal data (see Figure 2).

The conversion process attempts to ensure that only "usable" records are allowed on the database. This process also adds the various categories and other indicators (e.g., Major Diagnostic Category code, Service Types) that make the data more useful to the end-user for analysis. The Case build process interrogates inpatient medical claims and creates associations to all other claims related to the particular inpatient instance (e.g., related care to the hospitalization). This process allows the user to analyze patterns of treatment for illness that require in-patient hospitalization. The Episodes build process interrogates all types of claims and creates an episode for a particular illness (e.g., diabetes, asthma). This module allows concurrent episodes of unrelated illness to be associated in separate episodes.

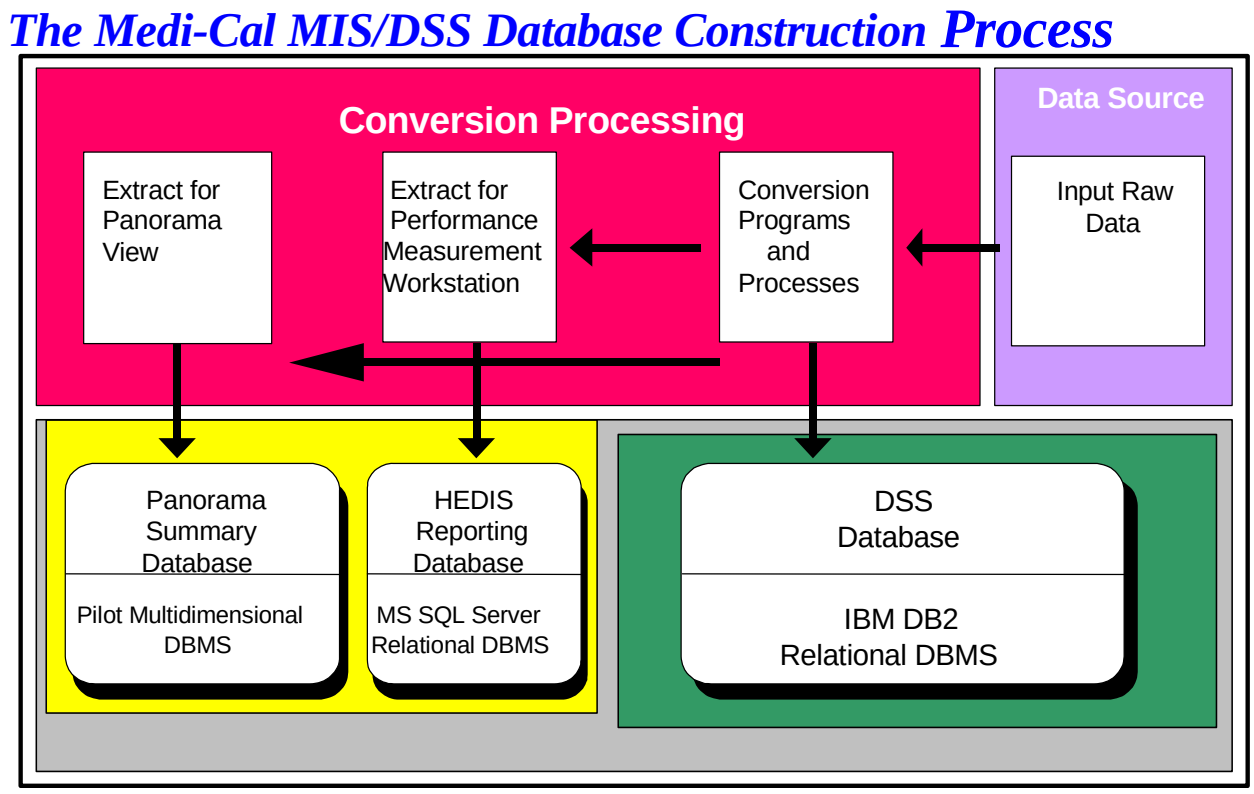

Figure 2. The MIS/DSS Database Construction Process

1.2 MIS/DSS Databases

1.2.1 DataScan®

The DataScan® DB2 database is a comprehensive data warehouse containing detail-level records of eligibility, medical services (fee for service and prepaid), drug prescriptions, providers, capitation payments and managed care plan financial reports. Its primary purpose is to allow analysis of detailed data.

1.2.2 PMW

The PMW SQL Server database also stores detail-level records, but it selects from the DataScan® warehouse only those records needed for HEDIS reporting. Its primary purpose is to create HEDIS reports that comply with NCQA requirements.

1.2.3 Panorama View®

The Panorama View® database contains highly aggregated data in a special database format designed for quick retrieval. The Pilot multidimensional database (MDDB) does not contain detail records, but pre-aggregated summary "cells". These cells hold values for common aggregations (e.g., by service category, provider type, and eligibility categories) that can be used to quickly answer questions for management about the overall trends of the expenditures, enrollment, quality, and access.

1.3 MIS/DSS Applications

1.3.1 DataScan

This product comprises a powerful, easy to use set of applications specifically designed to enable analysts to understand trends, identify patterns of care, and access detailed information on providers, patients, procedures, and diagnoses in the Data Warehouse.

1.3.2 Panorama View

Panorama View is based on an easy to use interface that users can interact with and ask and answer a variety of iterative questions based on a summary data set.

1.3.3 Panorama Briefing Book

This product provides a point-and-click display of preformatted, customized reports tailored to Medi-Cal requirements.

1.3.4 MyEureka!

MyEureka! supplements DataScan custom report capabilities and provides trained analysts with the user-friendly capability to design their own queries and reports from the detailed data warehouse.

1.3.5 Performance Measurement Workstation

This application calculates the HEDIS measures that can be supported through administrative data.

MEDSTAT supplements its products with powerful third party software that will provide additional analytical capabilities.

1.3.6 MapInfo

MapInfo is a PC-based geographic information system that supplements the mapping capabilities built into Panorama View.

1.3.7 SAS

SAS is the industry's leading statistical software designed to support complex statistical analyses requiring regression and/or correlation.

1.4 Database Updates

The DataScan databases contain 30 months worth of data, which are updated monthly, which results in the oldest of the 30 months "rolling off" to accommodate the new month. The Panorama View database includes 30 months of paid data but displays 27 months of data based on the date of service. The PMW database is updated annually and with each Phase of the project.

In the production environment cycle of the MIS/DSS, monthly updates of Claims, Eligibility, Provider, Capitation, and Managed Care Financial data will occur. These updates take approximately 17 working days after all source data is received. The process used to update the database occurs in the "background" and is transparent to users. As the process comes to completion, a "turnover" will occur in which the new month of data will become viewable while the oldest month in the thirty month database will "roll-off" (i.e., be deleted). The impact to the end-user is that constant consideration must be made to trending or comparison analysis when the database is about to be updated. For example, if a report were run one day and the update "turnover" occurred that evening, a different 30 months of data would be visible the following day (i.e., result would be different than the previous day).

2. Database Overview for DataScan Database

2.1 Background

The DataScan database is the foundation of the Medi-Cal Management Information System/Decision Support System (MIS/DSS). This database is the data warehouse, containing detail-level data for medical claims and encounters, prescription drugs, eligibility, providers, capitation payments and managed care plan financial reports. It also contains additional useful fields (such as Medi-Cal age group and Major Diagnostic Category on the Claim tables) at the detail level, plus aggregate-level information about inpatient cases and episodes of care.

Medi-Cal "raw" data is enhanced for loading into the DataScan database. Then the data for the Panorama View and Performance Measurement Workstation (PMW) is extracted from the DataScan database (or the table load files). So the DataScan database is important not only as the detail-level repository, but also as the source of the other databases in the MIS/DSS.

Because of its foundational role, the data enhancements (transformations) that produce the DataScan database are critical. Our System Design deliverable contains detailed descriptions of every custom data transformation for the DataScan database. This is supplemented by the Data Management Guide, the documentation that describes in detail the standard ("core") data enhancements.

The DataScan database, of course, was designed for the DataScan application. It is accessible by other applications, however, and has a broader role in the MIS/DSS than just supporting the DataScan application. For the most part, this database overview treats the DataScan database apart from the application.

2.2 The DataScan DSS database

2.2.1 Major tables

The DataScan database is implemented on IBM's DB2 relational database management system (RDBMS). The DataScan database comprises about 150 tables, of which about a dozen store the actual Medi-Cal data. The rest are support tables, so called because they support either the database build (validation, mapping, parameter and worktables) or the DataScan application (catalog, security, lookup, reporting, scripting, subsetting and norm tables).

Following is a brief description of each of the major Medi-Cal data tables. The tables have been grouped by content. See the Table Background and Field-Level Functional Specifications for each table for more detail.

2.2.1.1 Medical services and prescription drug tables

- Claim. These tables contain the detail-level records of medical services and encounters. There are seven of these tables, but for most purposes they are conceptually just one. The reasons for multiple tables have to do with performance, the relationship of the Inpatient table to the Case table, and the way we treat services that were paid within the 30-month database window but were performed prior to the window. The seven tables are:
  - Outpatient Claim 1 - 4. These tables contain all medical services and encounters which (1) are not part of a Case and (2) both the service and paid dates were within the 30-month database window. There are four tables to aid processing efficiency and query performance.
  - Note that "Outpatient" here means "not part of an Inpatient Case." Because of the way the Case build program works, there may be services/encounters on these tables, which occurred in an inpatient setting. In effect, "Inpatient" is short for "Inpatient Case detail," and "Outpatient" means "all other medical services/encounters."
  - Outpatient Paid. Records move to this table from the Outpatient Claim tables when their dates of service fall outside the 30-month database window, but their paid dates are still within the window.
  - Inpatient Claim. This contains all records, which belong to a Case, as defined by the Case table. All records on this table are inpatient, but not all inpatient services are necessarily on this table. Each service is linked to its Case by the CASEID field.
  - Inpatient Paid. A record moves to this table from the Inpatient Claim table when its Case-level admission date drops out of the 30-month window, but the record's paid date is still within the window. Unlike the Outpatient Paid table, records on this table can have service dates both within and before the 30-month window, since inclusion here is based on the Case-level admission date, not the service date of the detail service record.
- Drug. This table contains detail records for prescription drugs and medical supplies.
- Case. This is an aggregate table, containing summary records for inpatient cases (admissions, discharges). The DataScan Case build program uses MEDSTAT proprietary methodology to identify Cases and assign Case-level fields, including length of stay, DRG, MDC, disease stage and a number of financial fields. All the detail records, which make up the Case are on the Inpatient Claim table and are linked to the Case record by the CASEID field. DataScan Cases include both hospital and professional services.
- Episode. This is another aggregate table, containing summary records for episodes of care for individual patients. An Episode can include both inpatient Cases and Outpatient Claim records, and new with DataScan 4.2, Drug records. They are all linked to the Episode via the EPIID field. Each Episode record has Episode-level fields, such as disease classification and severest stage reached, number of inpatient Cases included and total inpatient days, and a range of financial summaries.

2.2.1.2 Eligibility tables

- Eligibility. This table stores one record per eligible per month. It contains such information about the eligible as CIN, birth date, eligibility category, ethnicity, etc. During the initial enhancement (conversion) of Claim and Drug data, certain fields are copied (tagged) from this table.
- Population. This is an aggregate table which stores counts of eligibles by month by a variety of variables (Medi-Cal age group, county, aid code and category, etc.). These counts are used as the denominators for rate calculations in various DataScan reports.
- Vital Statistics (DHS Core). This table contains the most recent data for certain fields for each eligible.

2.2.1.3 Provider tables

- Provider Directory. This table contains one record for each unique provider ID (PROVID). Its primary function is to supply provider names on DataScan reports.
- Provider Background. This table is not available to end-users. It is listed here because it is used to tag certain fields, including the unique provider ID and latitude and longitude codes, to the Claim and Drug records during their conversion.

2.2.1.4 Financial tables

- Capitation Payments. This table stores information about capitation payments to managed care plans. Fields include the payment amount, member months covered, PHP code, payment and service dates, county and aid code.
- Managed Care Plan Financials. This table stores data from “Orange Blank” forms that managed care plans are required to file quarterly with the California Department of Corporations.

2.2.2 Access

End users cannot access the DataScan tables directly. Rather, they access them through DB2 views, in keeping with standard database security practice. These views can be accessed via SAS, MyEureka! or any other program with DB2 connectivity, such as SPUFI (an IBM utility). In order to access a table, the database administrator (DBA) must have created a view on the table and assigned the user rights on the view.

See the DataScan Database Administrator’s Guide, chapter 6, for more details about security views.

2.2.3 Indexes

DB2 indexes on the major tables are designed to enhance performance. Some tables have multiple indexes. It is important for indexes to be designed to reflect end user subsets (selection or filter criteria), as this can greatly increase performance.

The primary index is clustering, which means that the records are physically stored in sorted order. The primary indexes for the major DataScan tables are:

2.2.3.1 Claim and Drug:

PRODUCT	Medical plan
NETWORK	Medical plan model type
ELIGCNTY	Eligible's county
EMPID	Client Identification Number

2.2.3.2 Case and Episode:

PRODUCT	Medical plan
ELIGCAT	Eligible's aid category
NETWORK	Medical plan model type
ELIGCNTY	Eligible's county
EMPID	Client Identification Number

2.2.3.3 Eligibility:

RELMO	A numeric representation of the relative month of eligibility in relation to the 34 months of data (30 months plus 4 new months).
ENROLLDT	Eligible's Enrollment Date
EMPID	Client Identification Number

2.2.3.4 Population:

EMPID	Client Identification Number

2.2.3.5 Vital Statistics:

EMPID Client Identification Number

2.2.3.6 Provider Directory:

PROVID Unique Provider ID

2.2.3.7 Provider Background:

PROVID Unique Provider ID

2.2.3.8 Capitation:

No index

2.2.3.9 Managed Care Plan Financials:

PLANMCF Corporate level PHP code

QTREND Last month in the quarter reported (CCYYMM format)

RPTNBR Report identifier

LINENBR Line number on the source report

2.2.4 Missing values

RDBMSs like DB2 support a special value called NULL, which means "unknown value." With a few exceptions, the DataScan database does not allow NULL values. This design characteristic arises from considerations of storage space (allowing a field to be NULL adds a byte to the field's storage length), and ease of SQL maintenance for the DataScan application and performance.

Rather than allowing NULLs, each field is assigned a specific value to designate "missing." For a character field, this is typically spaces. For financial fields it is usually zero, and for numeric categorical fields –1. These are *typical* values. The missing value is listed in the Field-Level Functional Specification for each DataScan field.

Actually, this use of a missing value can make some queries work more intuitively. Here is an example from the Eligibility table. Suppose that you want to count eligibles in Sacramento County. The filter criterion is ELIGCNTY = 34. If you want to count all eligibles outside Sacramento County, the filter criterion is ELIGCNTY <> 34. This count will include all

eligibles with an invalid (ELIGCNTY = 98) or unknown/missing county (ELIGCNTY = 0, the missing value). Thus the sum of the two counts equals the total number of all Eligibility records.

On the other hand, if you want to find all eligibles born, say, before 1996, you would probably use the filter criterion BIRTHDT < '1996-01-01'. You would get a misleading result, however, because the missing value for this field is the date January 1, 1 (0001-01-01), and your count would include everyone with an unknown date of birth. To get the expected results, your filter criterion would have to include AND BIRTHDT <> '0001-01-01'. Note that the DataScan application takes missing values into account, so a custom report defined within DataScan would not need the latter condition applied.

2.3 Applications that access the DataScan DSS database

2.3.1 DataScan

The DataScan DSS database was developed specifically to work with the DataScan application. This application divides the major tables into two types, core and custom. Core tables are required for every DataScan installation, and are used in standard reports. Custom tables are optional, and therefore are not used in any standard reports. Rather, they can be accessed only through the Record Listing and Custom Reports functions.

Similarly, fields are classified as core, managed care or custom. Core fields are required on every DataScan database, while managed care fields are required only for databases with managed care data. Custom fields are defined by the client. Custom fields are accessible only through Record Listing and Custom Reports.

The DataScan application maintains its own catalog of the DataScan DSS database. Among the functions of this catalog are to:

- Map PC field names used throughout the DataScan documentation to the DB2 column names. The DataScan application presents to the user the PC field names and short descriptions of their meaning rather than the DB2 column names.
- Define which fields users can see and which are hidden.
- Define table groups, which allow tables to be treated together in subsetting and Custom Reports operations.

The DataScan application uses several support tables. For example, it uses the Column Lookup table to assign labels on reports and to populate selection lists for report definitions. This table contains the valid values and their English descriptions for most categorical fields on the major tables. A categorical field is one with a defined set of valid codes, such as category of service, county, aid code or Medi-Cal age group.

2.3.2 MyEureka!

MyEureka! features graphical, point-and-click query and report definition capabilities. It accesses the DataScan DSS database via ODBC, a database connectivity method. Like the DataScan application, MyEureka! has its own catalog, called Metadata, to define the tables and columns of the DataScan database to which it has access. During the development phases of the MIS/DSS project, the Metadata file is maintained by MEDSTAT personnel.

2.3.3 SAS

SAS can be used for advanced statistical analyses. To access the DataScan DSS data, the user must use the DB2 view and column names, not the short names used throughout the DataScan database specifications and other documentation. The DB2 column names and data types are included on the Field-Level Functional Specifications.

2.3.4 Other Applications

Other user applications, which can access DB2 databases, can access the views in the DataScan DSS. While not part of the MEDSTAT-provided MIS/DSS application suite, these might prove useful to some users. An example is SPUFI, an IBM SQL (Structured Query Language) query tool, which MEDSTAT uses extensively in system testing. As with SAS, DB2 view and column names must be used in all SQL queries submitted with such tools.

2.4 Data Flow

Figure 3 provides a high-level view of the flow of data from “raw” data input to the various DataScan tables. This chart depicts the data flow rather than the process flow, which is much more convoluted.

Provider records from the Merged Provider Master File (PMF) and the Managed Care Plan (MCP) Provider File are consolidated and loaded into the Provider Background table. The consolidated data is aggregated on unique provider ID (PROVID) and loaded into the Provider Directory table. Since a single PROVID can represent more than one Medi-Cal ID, the Provider Directory represents the information for the most recently assigned active Medi-Cal ID.

MEDS data is transformed and loaded into the Eligibility and Vital Statistics tables. The Eligibility data is then aggregated into the Population table.

The RF-035 data is first split into Claim and Drug records, which are then transformed by their respective convert and edit programs. The loaded Provider Background and Eligibility tables are used to “tag” certain fields onto the Claim and Drug records. The edited Drug records are loaded into the Drug table. The Claim records actually take a more roundabout route to the Claim tables, but eventually they all get loaded into one of those seven tables. The edited claims also

go through the Case build, which produces the Case table. The Episode build uses the Case and Outpatient Claim records to construct and load the Episode table data.

Capitation data is transformed and loaded into the Capitation table.

Managed Care Plan Financial data is submitted on three input files, converted, then consolidated into the MCP Financials table.

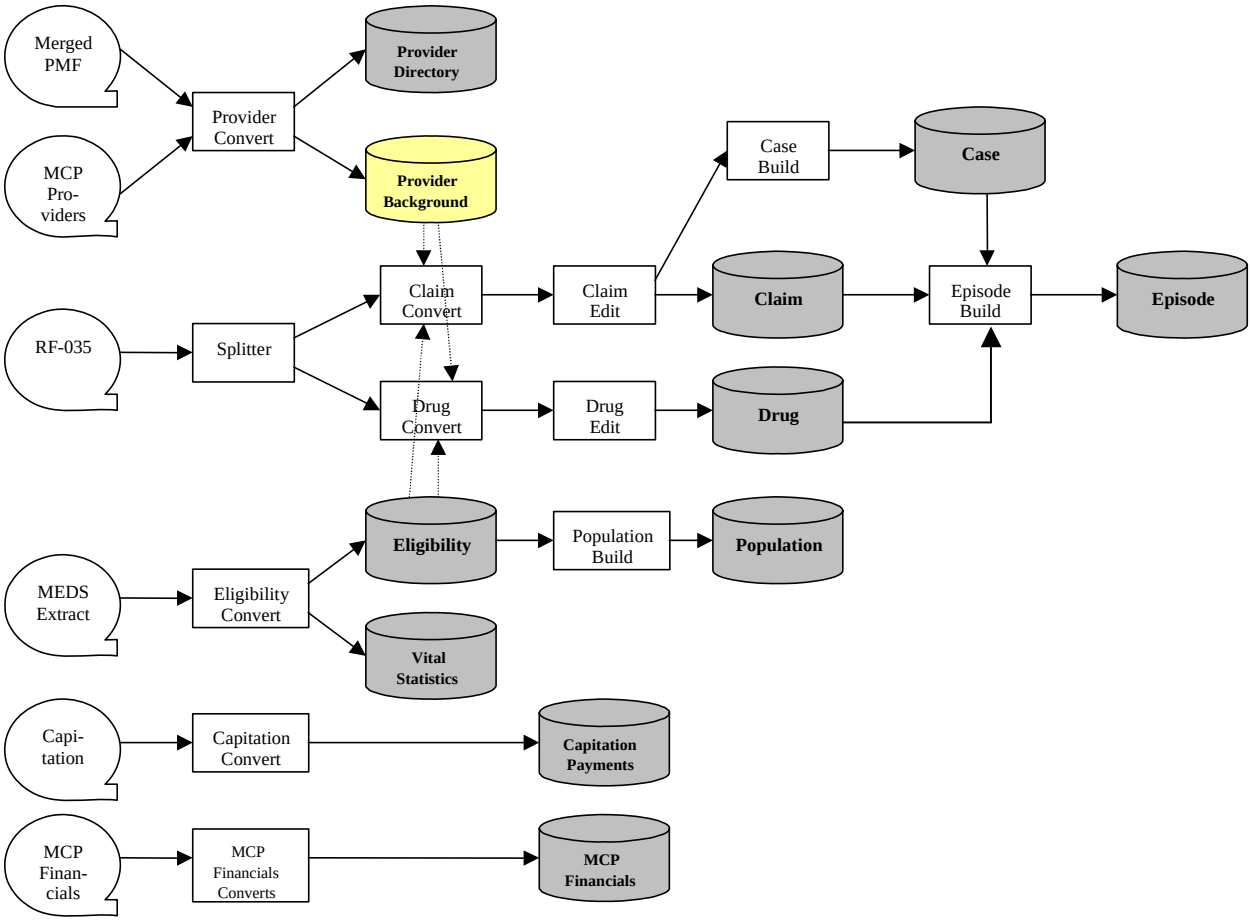

Figure 3. Data Flow

2.5 Data Enhancements

2.5.1 Installation

In an initial (phase) installation, 30 months of "paid" or processed data is included as input to the database. Note that the service date for these 30 months may occur in advance of the 30 month window – these services are included in the paid tables of the database and are therefore

available to end-users through the paid views within the DataScan application. All client data tables are loaded "from scratch" during the installation of each Phase.

2.5.2 Periodic updates

The DataScan tables are updated monthly. Essentially, a new month of paid data is added, and the oldest month of data on the database is deleted.

There are variations on this theme, which are discussed in the following paragraphs.

2.5.2.1 Claim, Case and Episode tables

All Cases are reconstructed for every *patient* (EMPID + MEMBERNO) who has at least one new caseable Claim record. Existing Episodes also are rebuilt if new Cases or Outpatient Claims can be added. Furthermore, some Claims roll off the database in two stages. First, if the paid date is within the new 30-month window but the service date is previous to the window, the Claim record moves to a Paid table. Then when the paid date falls out of the window in a later update, the record is deleted. The rebuilding of Cases and Episodes and the two-stage roll-off has side effects on existing Claim records, including:

- Some records may move from an Outpatient table to the Inpatient table, or vice versa.
- Some may be reassigned to different Cases or Episodes.
- Some may move to the Inpatient or Outpatient Paid table.

Because the update takes two weeks or longer to complete, a way must be found to keep the "old" database available for use during the update process without including any new data, then switching over cleanly to the "new" database. MEDSTAT does this by using the APPLIND field in its user access views. This field has three valid values during the update process. For example, assume an update will add the month of July 1999. During the update, there will actually be 31 months of data in the DataScan tables, although only 30 months are available for querying by the end user. This is controlled through "views" which use the APPLIND field. Prior to the update completing, APPLIND has the following values:

Y Data in both the "old" and the "new" window: February 1997 – June 1999.

N New data added by the update: July 1999.

D Data in the oldest month of the "old" 30-month window, to be deleted at the end of the update process: January 1997.

The user views contain all rows where APPLIND = Y or D (the "old" database). There is another set of views used for testing the new database before releasing it to the end users, which include rows where APPLIND = Y or N (the "new" database). At the end of the process, the D

rows are deleted, and all the Ns are changed to Ys. This prepares the database for the next monthly update.

This "update in place" process is used for the medical service, prescription drug, case, and episodes tables.

2.5.2.2 Drug Table

There are no special considerations for this table.

2.5.2.3 Eligibility Tables

In addition to the new month of data, we also process a new version of the most recent previous three months, which we call retroactive Eligibility data. This retroactive data replaces the data already in the database for those three months. The Population table is completely rebuilt every update.

2.5.2.4 Provider Tables

Information Technology Systems Division (ITSD) provides complete Merged PMF and MCP Provider files each month, but we do not rebuild the Provider Load File (PLF) from scratch. Rather, we add new providers and replace records for existing providers. This assures that providers who no longer appear on the Merged PMF and MCP Provider files will stay on the PLF.

2.5.2.5 Capitation Table

There are no special considerations for this table.

2.5.2.6 Managed Care Plan Financials Table

Not all plans have the same reporting quarters, and even those that do not all submit their reports at the same time. Therefore, the update data submitted will include reports from a variety of quarters.

2.6 Understanding the specifications

The complete set of data enhancement specifications comprises documents on four levels:

- System Overview
- Database Overview (this document)

- Table Background, one for each table or closely-related set of tables (e.g., Provider Directory and Provider Background tables are treated in a single document)
- Detailed specifications at the Field Level

The lowest two levels of documentation exist only for the DataScan database. Following are explanations of some of their features.

2.6.1 Table names and PC field names

As in this document, the table background and field-level specifications use PC field names rather than DB2 column names. Similarly, conventional table names (Claim, Eligibility, etc.) are used rather than DB2 table names. These conventions are followed throughout MEDSTAT's documentation.

In the field-level specifications, the DB2 column names and data types are also displayed for reference.

2.6.2 Enhancement Methods

See the System Overview for a discussion of the purposes of data enhancement. Here we discuss the methods. Fundamentally, there are two levels of data enhancement: record-level and multi-record.

In **record-level** enhancements, all the input fields are on the same record. Conversion programs transform Medi-Cal input data, outputting one record for each input record. Edit programs are similar, but their input is the output of a conversion program.

The following bullets explain MEDSTAT's most common record-level enhancement methods, and how they are presented in the specifications

- *Assign a constant*. Here the same value is assigned to a field in every record, regardless of input values. This is generally done only for DataScan required fields, which the Medi-Cal data does not support. For example, PCPID is always set to missing in the Claim and Drug tables, because the F35 file does not store a primary care physician's identifier.

  Example: EMPREL = 1.

- Move as is. This simply means that the output field contains the input field value unchanged. We almost always validate the data type, however, checking that output fields defined as CHARACTER in DB2 contain only alphanumeric data, and that numeric output fields (SMALLINT, INTEGER and DECIMAL DB2 data types) contain only numeric data. (Technical note: We treat low values as non-numeric.)

  Example: COPAY = F35-COPAY-AMOUNT (if numeric) **rounded to the nearest dollar**.

- *Validate*. This term is short for "validate and move as is, if the value is valid." We validate an input value by looking it up in a list of valid values. We typically call this list a map, even if the input values are not mapped to different output values. Our convention for validation (and move as is) is:

  Validate <Input field name> in <Map name>, matching on:

  <Input field name> = <Map name> <Map's key column name>.

  We provide maps and validation tables, annotated with the meanings of the key and lookup values, as part of the System Design deliverable.

  Example:

  Validate POPS-LANG-CODE in the LANGUAGE map, matching on:

  POPS-LANG-CODE = LANGUAGE Code.

  This specification references this map:

Medi-Cal Valid Values for LANGUAGE	
Code	Description
0	American Sign Language
1	Spanish
2	Cantonese
3	Japanese
4	Korean
...	...

  The result of this transformation is that if the value of POPS-LANG-CODE is in the key column (Code) of the map, that value is moved to the output field. If it is not in that valid value list, it is not moved to the output field.

- *Map*. An input value is transformed into an (usually but not necessarily different) output value via a map or lookup table. Not surprisingly, our specification convention is very similar to that for validation (see above).

  Look up [or Map] <*OUTPUT field name*> in [from] <*Map name*>, matching on:

  <*Input field name*> = <*Map name*> <*Map's key column name*>.

  We provide maps, annotated with the meanings of the key and lookup values, as part of the System Design deliverable.

  Example:

  Map ELIGCAT from the ELIGCAT map, matching on:

  AIDCODE = ELIGCAT Aid Code.

  This specification references this map:

Aid Category (ELIGCAT) Map from source AID CODE

Source Field: Aid Code	Description	MDST Aid Category (ELIGCAT)	Description
00	Invalid	99	Other / Invalid
01	Refugee Cash Assist	41	Refugee/Entrant
03	Adoption	21	Med Indigent Adults
...	...	...	...

  The result of this transformation is that if AIDCODE = 01, for example, then ELIGCAT = 41. If the AIDCODE value does not occur in the key column of the map, no value is assigned to ELIGCAT from the mapping operation.

- *Tag*. Tagging is a special type of mapping, where the lookup table is itself the output of a conversion process. For Medi-Cal, the Claim and Drug convert programs tag fields from the Eligibility and Provider Background tables. The specification format is essentially the same as for mapping.

  Example:

  Look up EMPZIP in (tag from) the Eligibility table, matching on:

  F35-CIN = EMPID

  SVCDATE = ENROLLDT (match on year and month only).

- *Calculate*. A field is the sum, difference, concatenation, or some other combination of two or more input fields.

  Example:

  PAY = COB + COPAY + DEDUCT + NETPAY.

- *Conditional logic*. Two or more instances of the above methods are applied conditionally. With the new functional specs we are avoiding the nested If statements of pseudo-code, since they are very hard to decipher. Rather, we use sets of rules, often ordered, to describe conditional logic.

  Example:

  Try to tag NETWORK from Eligibility. If that fails, map from PHPCODE.

  1. Look up NETWORK in (tag from) Eligibility, matching on:

     F35-CIN = Elig EMPID

     SVCDATE = Elig ENROLLDT (match only on year and month).

  2. If the tag fails to return a non-missing value, look up NETWORK in the NETPROD map, matching on:

     PHPCODE = NetProd PHP CODE.

  Sometimes the conditional logic is most easily presented in a logic table (not to be confused with a validation or mapping table). Wherever we use a table we explain how to read it. In general, you find the first rows in the table where all the input values or conditions are true, then use the output value in the last column of that row.

  To take advantage of Microsoft Word's table formatting capabilities, tables are placed in attachments to the specifications. The name of the attachment is the same as the output field.

  Example (for ADMITYP):

  ADMITYP = F35-ADMIT-TYPE, *except* for UB-92 claims (F35-CLAIM-FORM-INDICATOR = U). For the latter, the admission source (F35-ADMIT-SOURCE) must also be considered.

  The following table contains the rules for setting ADMITYP. To use the table, find the row, which contains the values of the three input fields. Assign ADMITYP to the value in the fourth column of that row.

  Only the listed combinations of the three input fields yield a non-missing value for ADMITYP. If none of the rules is satisfied, set ADMITYP to space, the missing value.

F35-CLAIM-FORM-INDICATOR	F35-ADMIT-SOURCE	F35-ADMIT-TYPE	Set ADMITYP to:
Space (Not UB-92)	Space	1 – 6	F35-ADMIT-TYPE
U (UB-92)	Space	1	1 (Emergency)
U	Space	3	2 (Elective)
U	Space	4	7 (Newborn)
U	4 (Emergency transfer) 5 (Elective transfer) 6 (Delivery transfer)	1	4 (Emergency – transfer)
U	4, 5, 6	3	5 (Elective – transfer)

In this example, suppose your input values are:

F35-CLAIM-FORM-INDICATOR: Space

F35-ADMIT-SOURCE: Space

F35-ADMIT-TYPE: 4

The output value is 4 (F35-ADMIT-TYPE), since all the input values match the conditions on the first row.

Now, suppose your input values are:

F35-CLAIM-FORM-INDICATOR: U

F35-ADMIT-SOURCE: 5

F35-ADMIT-TYPE: 3

This time the output value is 5, since the last row is the first one whose conditions the input values satisfy.

Finally, suppose your input values are:

F35-CLAIM-FORM-INDICATOR: U

F35-ADMIT-SOURCE: 5

F35-ADMIT-TYPE: 2

Since these input values match none of the rows on the table, no value is assigned from the table, and so ADMITYP defaults to the missing value.

In **multi-record** enhancements, data from two or more input records are combined into a single output record. Examples are Case build (combining input from several Claim records into a single inpatient Case), Episode build (combining input from Case, Outpatient Claim, and Drug records into a single Episode of Care) and Population build (aggregating Eligibility records into summary records with population counts by month).

In terms of how a specific output field relates to its input field(s) on the multiple records, there are three possibilities:

- Calculate the output value from the input values. Calculation methods include:
  - Sum, e.g., the Case field TOTCHG is the sum of CHG on all the Inpatient Claim records assigned to that Case.
  - Difference, e.g., the Case field LASTADM is the Case's ADMDATE minus the discharge date (ADMDATE + DAYS) of the most recent previous case for the same patient.
  - Count, e.g., the Episode field SRGCNT is the number of records assigned to the Episode with a procedure code classified as a surgical procedure.
  - Minimum, e.g., the Episode field FROMDATE is the earliest SVCDATE on the records assigned to the Episode.
  - Maximum, e.g., the Episode field STAGE is the highest value of STAGE on the records assigned to the Episode.
- *Move one of the input values.* This method is used when the summary record has a categorical field, e.g., AGE or LANGUAGE that also occurs on the records being summarized. Some rule must be used to decide which of the input values to move, since data problems or other situations can cause there to be more than one value represented. For example, a person may have an Episode whose date range includes the person's birthday, so that Outpatient Claim records assigned to that Episode will have two different values for AGE. For these kinds of fields DataScan typically used what we call the "demographic" rule, i.e., use the last non-missing value (in this case, the patient's age at the end of the Episode).
- *Use the input values to assign another value.* A good example is Diagnostic Related Group (DRG), which is assigned based on the diagnosis and procedure codes on the records in a Case.

All of these methods may be combined and involve conditional logic. Consider, for example, the Case fields PHYSID and PHYSNET. PHYSID is the PROVID from the Inpatient Claim records assigned to the Case, which has:

A PROVTYP of 50 – 99, but not 54, 77 or 84 (i.e., physicians, excluding anesthesiologists, radiologists and pathologists)

The highest total CHG of qualifying PROVIDs

This is a simplification; there are special rules when there are capitated services in the Case.

PHYSNET is the sum of NETPAY on all the Inpatient Claim records assigned to the Case for which PROVID = PHYSID.

2.6.3 Table overview format

Each of the major tables or groups of related tables above has its own specification, comprising a table overview document and a field-by-field detailed specification. The topics covered in the former are:

- Overview: How the table fits into the MIS/DSS system.
- Prerequisites/Pre-Conversion: What other MIS/DSS enhancement processes must be completed before this one can be accomplished.
- Special topics: Key enhancements for this table, such as unique provider ID for the Provider tables.
- Input data: What data sources feed into the table. File layouts are included as attachments.
- Output data: A listing of the outputs of the enhancement process.
- Process flow: A high-level view of how the data gets from the inputs to the final table(s).
- Installation: Special conditions or processing that occur only during a new database installation.
- Periodic updates: Special conditions or processing that occur only during a database update.
- Maps and validation tables: A listing of all custom maps that are used to validate and/or map values in the data enhancement.
- Tagging: What fields are tagged from the Eligibility and Provider Background tables.
- How to read the detailed (field-level) specifications: Special considerations for understanding the field-level specifications.

2.6.4 Field-level specification format

The field-level functional specifications are structured to logically present all the information needed to understand how an output field value is assigned. Refer to Figure 4 for an example while reading the following descriptions of the sections of the field-level specification.

Medi-Cal MIS/DSS
Data Enhancement Functional Specifications Field Level for Service Detail/Claim

Medical Service Claims

Output Field: **LATCODE** Latitude

Definition: Latitude assigned for use by MapInfo, consisting of a 7-digit signed number. Assigned based on the serving address of the billing provider.

DB2 Name: LATITUDE_MEASURE **Type:** Medi-Cal Custom Field **Missing Value:** None

Data Type: DECIMAL **Display Length:** 7, 4 **Storage Length:** 4

Input Fields:

Field Name	Data Type	Length
F35-PROVIDER-NUMBER	X	9

Logic: LATCODE is set during the tagging of PROVID. See field level spec for PROVID.

FOLOG Calls: OPR-#41: When tagging to the Provider File to find the latitude code of the Billing Provider (PROVID) and no match is found (call FOLOG with F35-PROVIDER-NUMBER).

Default: Zero . This will be assigned if the input value does not match *any* of the listed IDs in the Provider Background table.

Precedents: PROVID

Impact: Provider latitude and longitude codes are used with MapInfo to graphically display access patterns. For both Claim and Drug records, these reflect the location of the provider represented by PROVID. For Claim records, they reflect the servicing location of the billing ID.

Tech. Notes: A DB2 view has been created effective TB3.2 and thereafter to allow the display of 4 decimal places in LATCODE and LONGCODE fields in DataScan and IQ/Objects. This DB2 view allows the numeric fields to be viewed as character fields. When a field is viewed as a character field, no arithmetic operations can be performed on it.

In order for the field to be displayed as a character field in DataScan when it is stored as a decimal value, the DM Workbook will state that the field is decimal and has a data length of 4. The EGAD Detail will represent this field as a character field with a length of 10.

Revisions:

Date	Author(s)	Phase	IR(s)	Description
1/22/99	L. Richardson	4	1162	Indicated all the tagging logic resides underr PROVID.
1/21/99	C. Hubbert	4	1162	Clarified technical notes stating why there is a difference between the DM Workbook and the EGAD Detail.
12/28/98	L. Richardson	4	1162	Clarified the tagging logic.
11/16/98	C.Swanson	4	1162	Updated FOLOG with Operation Number and displayed field that corresponds to the Claims Convert Program.
10/28/98	C. Hubbert	3	824	Updated Technical Notes with DB2 view information
8/6/98	C. Hubbert	3	1000	Added PROVPLAN as one of the provider identifier fields to tag on.
8/5/98	C. Hubbert	3	1037	Add PROVLICU to match criteria for tagging. PROVLICU is the Uncollapsed Provider License number.
5/28/98	L. Richardson L. Macklem	3	824	Changed to numeric to support MAPINFO. Clarified that missing value is valid value.
5/18/98	L. Macklem	3	739	Rewrote as functional spec. In so doing, combined LATCODE and LONGCODE.

26-Jan-99 Page 1

Figure 4 Example Detail Field Level Functional Specification

- Output. This section lists for the output field:
  - Output Field: PC name and short description from the Data Element Dictionary (DED)
  - Definition: Full definition from the DED
  - DB2 Name: DataScan database DB2 column name
  - Type: How the DataScan application treats the field:
    - DataScan core field
    - Required for MC [managed care] reporting
    - Medi-Cal custom field
  - Missing value: The value that the DataScan application treats as missing
  - Data Type: DB2 data type
  - Display Length: The display length is the number of characters that displays on a DataScan report.
  - Storage Length: The DB2 storage length in bytes.
- Input field(s). This section lists each input field name, its data type and display length. An input field may be from the "raw" data source, e.g., the RF-035 file, or it may be an already converted output field. If the latter, it is also listed under Precedents below.

  A data type of **X** means the input field is alphanumeric, **9** means unsigned number and **S9** signed number.
- The next section describes in detail the data transformation for the field. We try to capture the what, why, and therefore of the field's data enhancement. Exactly, what transformation is occurring, why we do it that way, and what are the consequences.
  - Logic. This is the what, a description of the transformation, using the methods described above. It also includes statements as to why we do the transformation as we do. Although we try to write the logic in plain English, it should be as complete and unambiguous as the pseudo-code we used in Phases 1 and 2. And, of course, it should be much more understandable.
  - FOLOG Calls: A technical note describing the operation number and the criteria for the FOLOG call.
  - Default value. This is the value assigned to the output field if the Logic fails to assign a value.
  - Precedents. This is a list of all output fields which must be converted before the current logic can be executed, because those fields are inputs to this logic.

- ♦ Impact. This is the therefore, the consequences of the enhancement, including its impact of subsequent enhancements, the applications used with the database, the Panorama View and PMW extracts, and the user.
- ♦ Technical notes. These are notes that are useful, even necessary, for a comprehensive understanding of the transformation, but which would tend to clutter up the Logic subsection and make it harder to follow.

- Revision log. Each revision to the specification is listed in chronological order, most recent first. For each is listed the date, author(s), what phase of the project it was done for, the Investigation Request(s) (IR) initiating the change, and a description of the change.

2.7 Input data

Input data is provided by the Information Technology Systems Division (ITSD). During the development phases of the project, ITSD selects data appropriate to the phase. The following data tables are populated with primary data from ITSD, as stated:

- Claim: RF-035 file
- Drug: RF-035 file
- Eligibility: MEDS file extract
- Provider Directory and Background: Merged PMF and Managed Care Provider files
- Capitation Payments: The Managed Care Financial Unit's individual plan capitation statements, which DHS manually transforms specifically for the MIS/DSS
- Managed Care Plan Financials: The “Orange Blank” reports which managed care plans are required to submit to DHS quarterly

The following tables’ data is derived from one or more of the above tables, as follows:

- Case: Claim
- Episode: Case, Claim, and Drug
- Population: Eligibility
- Vital Statistics: Eligibility

2.8 Standard outputs

Typically a conversion program produces these significant outputs:

- *DataScan load file.* The most important output is a data file, which will be loaded into a DataScan DB2 table. Panorama View and PMW also use this data. Some "downstream" processes use the load file directly, while others extract their data from the DB2 tables.
- *Drop file.* This is a file containing all input records that were excluded from the output file(s). The counts of these records, by drop criterion, are given in the Aggregate Statistics report(s).
- *Aggregate Statistics report.* This report accounts for all input records by stating the number of:
  - ♦ Input records.
  - ♦ Records dropped (excluded), by reason.
  - ♦ Output records.

  Where appropriate, it also sums one or more financial fields for each of the items just stated. This report is important for tying the DataScan data back to the original input.
- *The FOLOG Report* documents records that have not been dropped but fail while converting raw input data into the format required for DataScan. The failure may be caused by one or more input fields that were not in the expected format (e.g., invalid data or non-numeric data in a numeric field). Please note that the FOLOG sometimes includes informational counts in addition to "failed" operations. For example, there is a FOLOG call to count the number of managed care encounters converted. The FOLOG Report includes:
  - ♦ Field name
  - ♦ Operation Number
  - ♦ Description of the operation that failed
  - ♦ Unmapped/undefined values found for that operation
  - ♦ Count of the number of records with possible errors for that operation
  - ♦ Percent of Total Records
  - ♦ Sum of the Net Payments
  - ♦ Percent of Total Net Payments

  The types of problems that the FOLOG report can highlight are:
  - ♦ Wrong input file was converted.

- Wrong conversion program was run against the input file.
- Input file format changed.
- Unmapped fields or field values were in the input data.
- Incoming input data values were all blanks or zeros.
- Unexpected field values were present in the input data.
- Improper records were dropped.

- *The Unexpected Values Report* is very similar to the FOLOG Report with several additions and will:
  - Indicate when a failed value has been confirmed by the State as an invalid value.
  - List the unmapped/undefined values found for each operation by PHPCODE.

The fields on the Unexpected Values Report are a subset of the FOLOG Report and are driven by two Excel spreadsheets. The first spreadsheet is a list of FOLOG operation numbers to be included in the report. The second is a list of previously approved values to map to other/invalid for each operation number. The State has the responsibility of determining which fields (only those listed in the FOLOG Report) to include in the Unexpected Values Report.

2.9 Maintaining the specifications

Beginning in Phase 2, all changes to the custom data enhancement specifications are managed by the change control process agreed to by DHS and MEDSTAT. The key document in this process is the Investigation Request (IR). All changes to the specifications will reference the appropriate IR number(s).

Beginning in Phase 3, MEDSTAT recast the specifications as functional specifications from pseudo-code. This effort will help DHS personnel who review the specifications and/or perform user acceptance testing to better understand the data transformations. This transformation to functional specifications was completed with the Phase 4 System Design Deliverable.

2.10 Related documentation

2.10.1 DataScan standard documentation

Of particular relevance are these two documents:

- Data Management Guide
- Database Administrator's Guide

Hard copies of these documents have been delivered to DHS. Beginning with DataScan 4.0, these documents are also available on CD.

2.10.2 Data Element Dictionary

The Data Element Dictionary (DED) lists for each field on the DataScan database's major data tables a description of the field's meaning and, where appropriate, a list of the valid values with their meanings.

2.10.3 IR Log

MEDSTAT maintains a log of all IRs. An IR may contain suggested changes to the enhancement logic, suggested database design changes, or just data investigation issues, which could lead to a change in logic and/or design. The IRs are reviewed regularly with DHS, and MEDSTAT delivers updated IRs to DHS.

3. Database Overview for Performance Measurement Workstation Database

3.1 Background

The MEDSTAT Performance Measurement Workstation 2000® (PMW 2000) will be used to build the Phase 5 database. This PMW database is designed to support comprehensive reporting of the National Committee on Quality Assurance's (NCQA's) HEDIS® 2000 reporting measurement set for the 1999 reporting year. Note that due to the timing, the 5.3 database will be built using PMW 2000 and the 1998 reporting year data. The 5.3.5 Monthly Update with June 2000 data will allow a rebuild of PMW using the 1999 reporting year data.

The MEDSTAT Performance Measurement Workstation 2000® (PMW 2000) database supports the HEDIS® reporting capabilities of the Medi-Cal Decision Support System (DSS). The PMW database is designed to support comprehensive reporting of the National Committee on Quality Assurance's (NCQA's) HEDIS® 2000 reporting measurement set. The fields contained in this database are specific to the HEDIS® 2000 requirements. The measures reported and the data that supports them is subject to revision whenever HEDIS® guidelines are changed. The HEDIS® 2000 guidelines impacted the following:

3.1.1 New measures for HEDIS® 2000 include:

- Chlamydia Screening in Women (First Year)
- Controlling High Blood Pressure (First Year)
- Use of Appropriate Medication for People with Asthma (First Year)

3.1.2 Retired Measures for HEDIS® 2000 include:

- Eye Exams for People with Diabetes is replaced by the Eye Exams numerator in the Comprehensive Diabetes Care measure

3.1.3 Domain Measures

The following list summarizes the complete set of HEDIS ® 2000 reporting measures now available in MEDSTAT Performance Measurement Workstation 2000®.

3.1.3.1 Effectiveness of Care (12 Measures)

- Childhood Immunization Status
- Adolescent Immunization Status
- Breast Cancer Screening
- Cervical Cancer Screening
- Chlamydia Screening in Women (First Year)
- Prenatal Care in the First Trimester
- Check-Ups After Delivery
- Controlling High Blood Pressure (First Year)
- Beta Blocker Treatment After a Heart Attack
- Cholesterol Management After Acute Cardiovascular Events
- Comprehensive Diabetes Care
- Use of Appropriate Medication for People with Asthma (First Year)
- Follow-Up After Hospitalization for Mental Illness
- Antidepressant Medication Management

3.1.3.2 Access/Availability of Care (5 Measures)

- Adults' Access to Preventive/Ambulatory Health Services
- Children's Access to Primary Care Providers
- Initiation of Prenatal Care
- Annual Dental Visit

- Availability of Language Interpretation Services

3.1.3.3 Health Plan Stability (4 Measures)

- Disenrollment
- Practitioner Turnover
- Years in Business/Total Membership
- Indicators of Financial Stability

3.1.3.4 Use of Services (17 Measures)

- Frequency of Ongoing Prenatal Care
- Well-Child Visits in the First 15 Months of Life
- Well-Child Visits in the Third, Fourth, Fifth and Sixth Year of Life
- Adolescent Well-Care Visits
- Frequency of Selected Procedures
- Inpatient Utilization - General Hospital/Acute Care
- Ambulatory Care.
- Inpatient Utilization: Non-Acute Care.
- Discharge and Average Length of Stay: Maternity Care
- Cesarean Section Rate
- Vaginal Birth After Cesarean Section Rate
- Births and Average Length of Stay: Newborns
- Mental Health Utilization: Inpatient Discharges and Average Length of Stay
- Mental Health Utilization: Percentage of Members Receiving Inpatient, Day/Night Care, and Ambulatory Services
- Chemical Dependency Utilization: Inpatient Discharges and Average Length of Stay
- Chemical Dependency Utilization: Percentage of Members Receiving Inpatient, Day/Night Care, and Ambulatory Services
- Outpatient Drug Utilization

3.1.3.5 Cost of Care (2 Measures)

- Rate Trends
- High-Occurrence/High-Cost DRGs

3.1.3.6 Health Plan Descriptive Information (8 Measures)

- Board Certification/Residency Completion
- Provider Compensation
- Arrangements with Public Health, Educational and Social Service Organizations
- Total Enrollment
- Enrollment by Payer
- Unduplicated Count of Medicaid Members
- Diversity of Medicaid Membership
- Weeks of Pregnancy at Time of Enrollment in the Health Plan

In order to produce the HEDIS® 2000 reports for Medi-Cal, the data is extracted from the Medi-Cal DataScan® database (or the table load files). The data is imported into the SQL Server database through ASCII flat files, conforming to a specific layout with specific data formats. Only the required records for a particular year of HEDIS® reporting are extracted, including any data that has been archived through the HEDIS archive feature of DataScan v4.2. The extract for Medi-Cal is done to include a six-month run-off of claims for a given reporting year before a year is reported via the PMW application to ensure maximum data availability for measures reporting.

3.2 Inputs

Input data to the PMW application is extracted from the DataScan® Database. The DataScan tables that are used for extract into the PMW build process include:

DataScan Table	Type of Data Extracted
Eligibility	Enrollment information
Inpatient Service	Inpatient services
Case	Inpatient services

DataScan Table	Type of Data Extracted
Outpatient Service	Services
Outpatient Paid	Services
Drug	Drug services
EGAD	Field interpretation
Col_Lookup	Custom field descriptions

Medi-Cal specifications for all PMW support tables are defined in further detail in the PMW Data Enhancement Functional Specifications Document.

3.3 Data Flow

Figure 5 gives a high-level view of the flow of data from "raw" data input to the various PMW tables. This chart depicts the data flow rather than the detailed process flow.

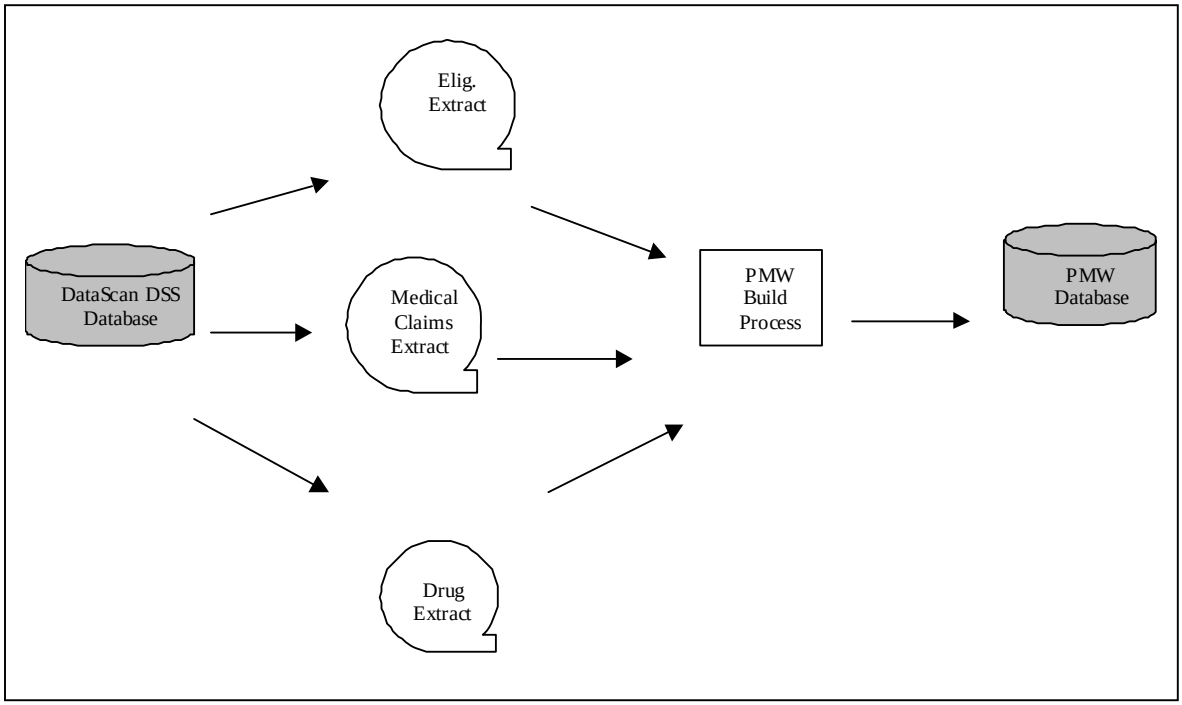

Figure 5. PMW Data Flow

3.4 Enhancement Processing

3.4.1 Installation

In an initial (phase) installation, we extract from 30 months of paid data to gain one complete HEDIS® reporting year of data. All client data tables are loaded "from scratch" during the installation. Enhancements performed during the extract and build process include clinical recoding, the collapsing of eligibility and the calculation of HEDIS® indicators.

3.4.2 Update Processing

The PMW Database is *always* reloaded. There is no update process for the PMW tables.

3.5 Outputs - The PMW DSS Database

The PMW database is implemented on Microsoft's SQL Server relational database management system (RDBMS). The PMW database is comprised of tables that store Membership data, Claims data, and Custom data.

The primary PMW tables are:

- Enrollment_History

  This table contains the enrollment (for Medi-Cal eligibility) records in a collapsed format (i.e., only enough records are stored to indicate coverage for the 30-month period) for up to two calendar years (if available). The fields included are only those necessary for reporting (e.g., ID number, coverage dates, aid codes, county, etc.).

- Patient_Profile

  This table includes data that uniquely identifies each patient that is in the Enrollment_History table. This data is derived from the DataScan Eligibility table used during the extract process. There is one record for every eligible that is in the Enrollment_History table.

- Op_Claim_Detail

  This table includes the outpatient services extracted from the DataScan Outpatient Services tables. The fields included are those necessary for reporting.

- Ip_Claim_Detail

  This table is where the inpatient services that were extracted from the DataScan Inpatient Service and Case tables reside. Like the Op_Claim_Detail table, the fields included are only those fields necessary for reporting HEDIS measures.

- Drug_Detail

This table contains the drug claim records originally extracted from the DataScan Drug table. Only those fields required for the reporting of HEDIS measures that require drug utilization related services are stored in this table.

- MCO

  This table contains the descriptions for all Managed Care Organizations (NETWORK) extracted from the DataScan Eligibility Table.

- Custom_1

  This table contains the descriptions of the first PMW custom field. In Medi-Cal, the descriptions for Eligible County (ELIGCNTY) are extract from the DataScan Col-_Lookup table and stored in the Custom_1_Desc.

- Custom_2

  This table contains descriptions for the second PMW custom field. For Medi-Cal, the descriptions for Aid Category (ELIGCAT) are extract from the DataScan Col_Lookup table and stored in the Custom_2_Desc table.

3.6 The Performance Measurement Workstation Application

The PMW DSS database was developed specifically to work with the PMW application. The application is used by Medi-Cal to generate HEDIS® reports for individual Health Plans, as well as for Fee-For-Service, Counties and eligibility categories.

3.7 Maintaining the specifications

Beginning in Phase 2, all changes to the custom data enhancement specifications are managed by the change control process agreed to by DHS and MEDSTAT. The key document in this process is the Investigation Request (IR). All changes to the specifications will reference the appropriate IR number(s).

Beginning in Phase 3, MEDSTAT recast the specifications as functional specifications from pseudo-code. This is to help DHS personnel who review the specifications and/or perform user acceptance testing, as well as end users, better understand what data transformations are being performed. The PMW Specifications Document in the System Design Deliverable includes all the functional specifications used to build the PMW database according to the Medi-Cal customization requirements. This transformation to functional specifications was completed with the Phase 4 System Design Deliverable.

3.8 Related documentation

3.8.1 PMW standard documentation

Of particular relevance are these three documents:

- PMW Technical Users Guide
- PMW Electronic User Guide within the PMW Help

4. Database Overview for Panorama View Database

4.1 Background

Panorama View® is the MIS portion of the Medi-Cal Management Information System/Decision Support System (MIS/DSS). This is a Medicaid-specific application that uses data, which has been summarized from the DataScan® DSS. Panorama View® provides quick answers to questions with a variety of graphical and tabular displays. It summarizes data at the high- and mid-level. It does not contain line-item detail like the DSS.

4.2 The Panorama View® MIS database

Relational database management systems (RDBMSs) like DB2 are optimized for transaction systems. They excel at efficient storage of large numbers of detail records and retrieval of *individual* records. They are not optimized for quick retrieval of summarized data. Multidimensional databases (MDDBs), on the other hand, are optimized for summary reporting and on-line analytic processing. Data is *pre-aggregated* in ways that users will commonly query it. Whereas an RDBMS favors flexibility over performance, an MDDB opts for speed over ad hoc flexibility.

The Panorama View® database is implemented using Pilot's Lightship MDDB, which structures the data by measures and dimensions. A *measure* is a value, such as total payments, count of eligibles or office visits/1000. A *dimension* is a category, such as county, aid category or provider specialty. Time is a special dimension.

4.2.1 Summary data cubes

The MDDB organization of data is less familiar than the row-and-column (or record-and-field) structure of relational database tables as in the DataScan® database. Rather than structuring the data in tables, Panorama View®'s Pilot MDDB structures data in three *cubes*: Cost/Use,

Provider and Quality. Rather than rows and columns, which limits the model to two dimensions, the data is stored by combinations of several dimensions, and each cell represents a measure at the intersection of a particular value for each dimension.

4.2.2 Measures

The Panorama View® database contains over 400 measures of healthcare cost, utilization, access and quality. Measures of cost and utilization are calculated based on three DataScan® database fields:

- SVCCAT (Federal category of service)
- NETPAY (Medi-Cal paid amount)
- UNITS (number of services)

In addition, fields supporting selection criteria (filters) for certain measures are captured from the DataScan® database records, such as diagnosis and procedure codes (DX1, DX2 ,PROC1, and AMBPROC).

4.2.3 Dimensions

The selection and proper treatment of dimensions is critical, because Panorama View® allows the user to drill down and subset on one or a combination of dimensions. For example, total expenditures can be analyzed by county, or by county and aid category (drill down). The user can limit the analysis to selected counties, age groups and aid categories, e.g., children in Sacramento County qualifying for Public Assistance to Families (subset). Dimensions can also roll up to a higher level, such as county to region.

Because Panorama View® data is extracted from the DataScan® database data files, most of the transformation logic which converts Medi-Cal "raw" input data fields into the correct values for Panorama dimensions occurs in the DataScan® database enhancements.

There are three (sets of) dimensions: Recipient, Provider and Time.

4.2.3.1 Recipient dimensions and roll-ups

Dimension	**DataScan® field**	**Summary Category Roll-Up (None if blank)**
Recipient Age Group	MCALAGE	Age Group Roll-Up
Recipient Gender	SEX	
Recipient Ethnicity	ETHNCTY	Ethnicity Roll-Up

Dimension	DataScan® field	Summary Category Roll-Up (None if blank)
Recipient County	ELIGCNTY	
Eligibility Category	ELIGCAT	Aid Category Roll-Up
Language	LANGUAGE	Language Roll-Up
Medical Plan	NETWORK	Medical Plan Model Type

4.2.3.2 Provider dimensions and roll-ups

Dimension	DataScan® field	Summary Category Roll-Up (Note: None defined)
Provider Type (Vendor)	VENDORCD	
Provider Specialty	PROVSPEC	
Provider County	PROVCNTY	

4.2.3.3 Time dimension and roll-ups

For the Cost/Use and Provider cubes, the time dimension for Claim and Drug data is service month (the year and month of SVCDATE) and for Eligibility data, the time dimension is enrollment date (ENROLLDT). The State has selected a 3-month lag to account for services not yet reported; therefore, 27 months of data will display in the Panorama View product for any 30-month time period. Months roll up into complete fiscal years and fiscal years-to-date. Because the underlying data covers 30 months, sometimes only one complete fiscal year of data is available by service date. There are always two years-to-date available.

For the Quality cube, the underlying time dimension is month of service (the year and month of SVCDATE), but the display is only by rolling year. The most recent six months of data are excluded, because of payment lag. The oldest 24 months on the database are presented as two consecutive years of data. The begin and end dates for the years change with each update.

4.2.4 Service Date Basis

With the release of Panorama View® version 2.0, the time dimension for all cubes uses the date of service. This service date view requires Completion Factors to be applied to the data to estimate claim payments and services that have not been reported – commonly known as Claims Lag. Panorama View computes an estimate to "complete" the data based on Medi-Cal's past claim lag experience. An Overview of the Completion Factor calculation and application can be found in Attachment 5 of the Panorama View Catalog Implementation Guide for Medi-Cal.

The Completion Factors are calculated and applied during the initial database build. During an Update, however, the factors are calculated, but not necessarily applied. The Completion Factors remain relatively stable over time (due to the consistency of claims lag and payments) and will therefore only need to be replaced on an annual basis. Beginning with the Phase 5 Updates, the State will be provided with reports comparing the Completion Factors calculated during the Update vs. the Completion Factors calculated during the initial 5.3 build. The State can request, via the Change Control process, to have the completion factors applied in a future monthly update.

4.3 The Panorama View® MIS application

Panorama View® belongs to the genus executive information system (EIS). It is designed to give a high-level view to decision-makers to help them see trends and patterns. It is structured as a set of questions organized into folders. The answer to each question is presented as a line graph, bar graph, map, or whatever graphical presentation is most appropriate. A tabular view of the data is also available. The application features an easy-to-use point-and-click user interface.

Medi-Cal has helped specify the questions and how the answers (measures) are calculated. Mid-level analysis is available through subsetting and drill down to the dimension level.

Note: The Panorama View® application is the only tool for accessing the Panorama View® database.

4.4 Data Flow

Figure 6 presents a high-level view of the data flow through the Panorama View® extract and enhancement process. The Quality cube is built from just Claim and Eligibility data. The Cost/Use and Provider cubes, however, include also Drug and Capitation data, plus bed counts from the DHS Licensing & Certification Facilities Directory.

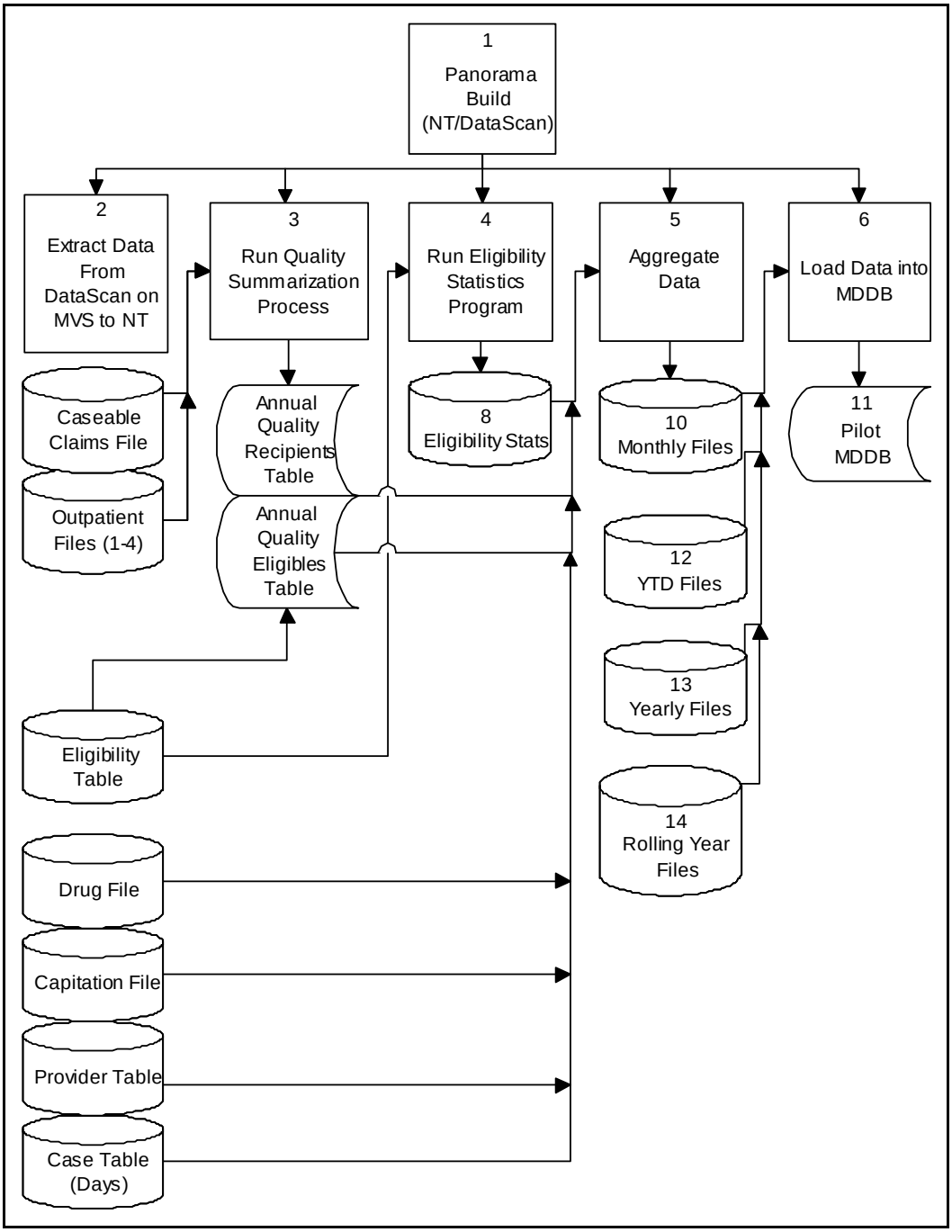

Figure 6 - Panorama View Extract and Enhancement Process

4.5 Data Enhancements

4.5.1 Installation

When a new database is installed for a phase, the entire 30 months of data is extracted for the Panorama View® database.

4.5.2 Periodic updates

Beginning with Panorama View® version 2.0, a full refresh of the MDDB tables is required because different completion factors are applied to specific months of service with each update cycle. At each update, new data are added incrementally to base tables on a date of service basis. This will include new months of data as well as additional data for months already in the base tables. (Base tables to be updated include data for recent months that were excluded in the previous update to the MDDB).

4.5.2.1 Claim and Drug

During a monthly update, the DataScan® database adds a month of paid Claim and Drug data. This data can, however, contain service dates from a number of months.

4.5.2.2 Case Days

During a monthly update, all existing Case Days records on the base table are deleted and replaced with an extract from the updated DataScan® Case Table.

4.5.2.3 Eligibility

As in the DataScan® database, newly submitted Eligibility data for the three most recent months as well as the new update month is incorporated in the rebuilding of the cubes.

4.5.2.4 Provider

The rebuild incorporates updated active provider (defined as one with claims or capitated services in the DataScan® Claim and Drug tables) data for all the months in the rebuild date range.

4.5.2.5 Capitation

Only the new month's capitation payments are distributed, since there is no retroactive capitation data.

4.6 Understanding The Specifications

Detailed specifications for the Panorama View® data enhancements are contained in the Panorama View® Catalog Implementation Guide. Like the Panorama View® application, the Implementation Guide is structured as a set of questions. The answers to these questions comprise the detail-level specifications for the Panorama View® database.

4.7 Input data

4.7.1 DataScan® data files

Data enhanced for the DataScan® database is input to the Panorama View® extract processes. In some cases this data is ready for loading into the DataScan® database; in other cases the data is extracted for the Panorama View database after it has been fully enhanced and loaded in the DataScan® database.

- Claim (medical services) data is extracted after the Claim edit program. At this point all record-level fields have been assigned, and the records have been divided into Caseable Claim and Outpatient files. Most of the Panorama View® Provider data also is extracted from these files.
- Drug data is extracted immediately after the convert program, before the DataScan® Drug edit program. No NDC-related fields are assigned at this point, but all fields required by the Panorama View® build are populated.
- Eligibility data is extracted from the file, which is loaded into the DataScan® Eligibility table. This is the same file from which the DataScan® Population table is aggregated.
- Active providers are defined as those, which have claims or encounters, so data other than bed counts is extracted from the Claim and Drug files as described above. Data from the DataScan® Provider Directory is not included in Panorama View®.
- Capitation payments data is extracted from the converted Capitation file. This is the file that is loaded into the DataScan® Capitation table.
- The Inpatient Case table is extracted immediately after the case construct process completes. This information is used to ensure that acute care days reported in Panorama View tie to those reported in the DataScan application.

4.7.2 U.S. Census Bureau data

Census data is downloaded from the U.S. Census Bureau site on the Internet World Wide Web.

4.7.3 DHS Licensing & Certification Facilities Directory

This file provides information for the bed counts in the Provider Access Folder.

4.8 Outputs

- MDDB load files. The main output is a set of files to load into the Panorama View® MDDB.
- "Skinny" table. An intermediary output is a "skinny" Claim and Drug table. This preserves the detail records for re-aggregating for the next monthly update.
- Log Files. As the build process runs, errors and other information such as record counts, payment amounts, etc. are logged into files. These log files are used to perform balancing and verification.

4.9 Maintaining the specifications

That the Panorama View® database is built from the DataScan® data files does not mean that it is just an aggregation of DataScan® data. The Panorama View® Extract, Summarize and Aggregate processes all entail a number of enhancements, from simple calculations to complex conditionally defined measures. Major enhancement methods are:

- Aggregate on dimensions. In the characteristic Panorama View® database build transformation, a measure is calculated as the sum of NETPAY or UNITS by each combination of dimensions. For example, calculate expenditures (sum of NETPAY) for each unique combination of MCALAGE, SEX, ETHNCTY, ELIGCNTY, ELIGCAT, LANGUAGE, NETWORK and SVCDATE year-and-month. This can produce thousands of values to be stored in various cells in the MDDB.

  Many measures use selection criteria. For example, each measure in the Quality cube selects records to aggregate based on specific diagnosis and procedure codes.
- Aggregate on dimension roll-ups. These further aggregate the dimension aggregates on the dimension roll-ups.
- Select unique values. For example, the Provider cube's count of active providers is based on the selection of unique combinations of PROVID, PROVCNTY, PROVSPEC and VENDORCD from the DataScan® Claim and Drug data files.
- Distribute (spread) values. The simplest case is the assignment of BEDCOUNT to facilities in the Provider cube from the DHS Licensing & Certification Facilities Directory. A much more complex case is the spreading of Capitation payments. The input data directly supports only four Recipient dimensions, and the data must be spread over the remaining dimensions based on eligibles' proportional distribution in the database. Special handling is performed on Medical Plans with confidential rate information.

- Convert from raw source. Some data does not come already converted and edited from the DataScan® data files. For example, census data is downloaded from the U.S. Census Bureau web site, and the Panorama View® database build must do whatever transformations and edits are required.

These enhancements are defined in a set of *catalogs*. These are parameters, rules and definitions of measures stored in Microsoft Access tables and text (.INI) files. All the Medi-Cal customizations for Panorama View® are defined in these catalogs, which are maintained by a specialized team in Ann Arbor, based on the Implementation Guide.

Beginning in Phase 2, all changes to the custom data enhancement specifications are managed by the change control process agreed to by DHS and MEDSTAT. The key document in this process is the Investigation Request (IR). All changes to the specifications will reference the appropriate IR number(s).

5. Summary of Document Changes

Date	Author	Phase	IRs	Description of Changes
3/22/00	**T. Poyner**	**5**	**1474**	**Added a line from the Drug table to the Episode Build in Figure 3.**
3/22/00	**C. Swanson**	**5**	**N/A**	**Updated according to HEDIS 2000 Specs.**
8/31/99	J. Dittman	5	N/A	Updated document with Phase 5 changes, specifically DataScan 4.2.
7/30/99	K. Key	5	1469	Changed all references to IQ/Objects to MyEureka!, the product's new name.
6/22/99	K. Key	4	N/A	Updated Indexes Section (2.2.3) to reflect the current index structure. Modified Claim/Drug, Eligibility, and Provider Directory.
6/16/99	J. Dittman	4	N/A	Updated document for P4 UAT Config. Manual with changes approved after the P4 System Design Deliverable.
1/25/99	T. Calvert	4	N/A	Consolidated the System Overview, DataScan Database Overview, Panorama View Database Overview, and Performance Measurement Workstation Overview documents into one document.

Date	Author	Phase	IRs	Description of Changes
1/25/99	J. Dittman	4		Updated document for Panorama View® version 2.0.
1/22/99	L. Richardson	4		Revised and reorganized the PMW document to offer increased clarity and to bring the format more closely in line with format being used in the Phase 4 System Design Deliverable. Revised the version numbers currently in effect.
6/8/98	L. Macklem	3		Created initial Panorama View document.
6/5/98	J. Mulcahy	3		Created initial PMW document.